



(19) RU (11) 2 181 481 (13) C2
(51) МПК⁷ G 10 L 19/12

РОССИЙСКОЕ АГЕНТСТВО
ПО ПАТЕНТАМ И ТОВАРНЫМ ЗНАКАМ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ РОССИЙСКОЙ ФЕДЕРАЦИИ

(21), (22) Заявка: 98101107/28, 13.06.1996
(24) Дата начала действия патента: 13.06.1996
(30) Приоритет: 16.06.1995 GB 9512284.2
(43) Дата публикации заявки: 27.11.1999
(46) Дата публикации: 20.04.2002
(56) Ссылки: WO 94/25959 A1, 10.11.1994. DE 3041423 C1, 16.04.1987. US 4617676, 14.10.1986. ПРОХОРОВ Ю.Н. Статистические модели и рекуррентное предсказание речевых сигналов, - М.: 1984, с. 7, 35, 88, 149, 161, 162, 168 и 181. GB 2258978 A, 24.02.1993. GB 2254986 A, 21.10.1992. GB 2235354 A, 27.02.1991. GB 2195220 A, 30.03.1988. RU 2140671 C1, 27.10.1999. RU 2007763 C1, 15.02.1994.

(85) Дата перевода заявки РСТ на национальную фазу: 16.01.1998

(86) Заявка РСТ:
GB 96/01428 (13.06.1996)

(87) Публикация РСТ:
WO 97/00516 (03.01.1997)

(98) Адрес для переписки:
105023, Москва, ул. Большая Семеновская 49,
оф.404, ООО Центр "ИННОТЭК", Аргасову О.В.

(71) Заявитель:
НОКИА МОБИЛ ФОУНС ЛИМИТЕД (FI)

(72) Изобретатель: ЯРВИНЕН Кари (FI),
ХОНКАНЕН Теро (FI)

(73) Патентообладатель:
НОКИА МОБИЛ ФОУНС ЛИМИТЕД (FI)

(74) Патентный поверенный:
Аргасов Олег Вячеславович

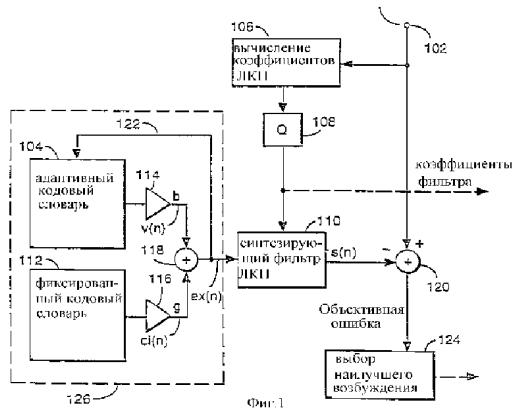
R
U
2
1
8
1
4
8
1
C
2

R
U
2
1
8
1
4
8
1
C
2

(54) СИНТЕЗАТОР И СПОСОБ ДЛЯ РЕЧЕВОГО СИНТЕЗА (ВАРИАНТЫ) И РАДИОУСТРОЙСТВО

(57) Изобретение относится к звуковому или речевому синтезатору для использования со сжатыми закодированными в цифровом виде звуковыми или речевыми сигналами и может быть использовано для постпроцессорной обработки сигналов, выделенных из словаря кодов возбуждения и словаря адаптивных кодов речевого декодера типа линейного кодирования с предсказанием (ЛКП). Постпроцессор работает по сигналу, выделенному из источника возбуждения,

путем добавления к нему масштабированного сигнала, выделенного из адаптивного кодового словаря. Коэффициент масштабирования определяется речевыми коэффициентами, вводимыми в генератор возбуждения. Полученный сигнал нормируется и вводится в декодер типа ЛКП или речевой синтезирующий фильтр перед подачей на блок звуковой обработки. Это позволяет улучшить воспринимаемое качество синтезированной речи. 8 с. и 31 з.п. ф-лы, 7 ил.





(19) RU (11) 2 181 481 (13) C2
(51) Int. Cl. 7 G 10 L 19/12

RUSSIAN AGENCY
FOR PATENTS AND TRADEMARKS

(12) ABSTRACT OF INVENTION

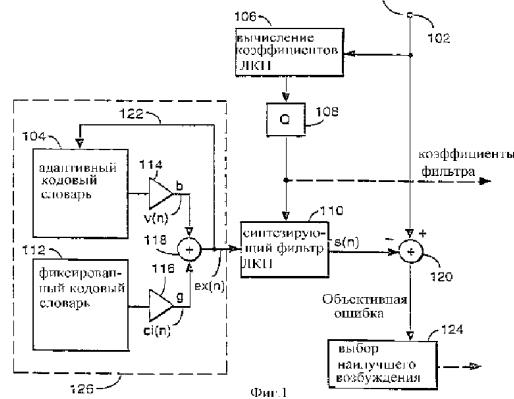
(21), (22) Application: 98101107/28, 13.06.1996
(24) Effective date for property rights: 13.06.1996
(30) Priority: 16.06.1995 GB 9512284.2
(43) Application published: 27.11.1999
(46) Date of publication: 20.04.2002
(85) Commencement of national phase: 16.01.1998
(86) PCT application:
GB 96/01428 (13.06.1996)
(87) PCT publication:
WO 97/00516 (03.01.1997)
(98) Mail address:
105023, Moskva, ul. Bol'shaja Semenovskaja
49, of.404, OOO Tsentr "INNOTEhK", Argasovu
O.V.

(71) Applicant:
NOKIA MOBIL FOUNS LIMITED (FI)
(72) Inventor: JaRVINEN Kari (FI),
KhONKANEN Tero (FI)
(73) Proprietor:
NOKIA MOBIL FOUNS LIMITED (FI)
(74) Representative:
Argasov Oleg Vjacheslavovich

(54) SYNTHESIZER AND METHOD OF SPEECH SYNTHESIS (VARIANTS) AND RADIO DEVICE

(57) Abstract:
FIELD: sound and speech synthesis.
SUBSTANCE: invention is related to sound or speech synthesizer used with compressed sound and speech signals encoded in digital form. It can be employed for post-processor processing of signals extracted from excitation code dictionary and adaptive code dictionary of speech decoder of linear predictive coding type. Post-processor is activated by signal extracted from excitation source by addition to it of scaled signal extracted from adaptive coding dictionary. Scaling factor is determined by speech coefficients entered into excitation generator. Obtained signal is set to norm and is input into decoder of linear predictive type or into synthesizing filter before it is supplied to unit of speech

processing. EFFECT: improved receiving quality of synthesized speech. 39 cl, 7 dwg



RU 2 181 481 C2

RU 2 181 481 C2

Изобретение относится к звуковому или речевому синтезатору для использования со сжатыми закодированными в цифровом виде звуковыми или речевыми сигналами. В частности, оно относится к пост-процессору для обработки сигналов, выделенных из словаря кодов возбуждения и словаря адаптивных кодов речевого декодера типа линейного кодирования с предсказанием (ЛКП).

В цифровых радиотелефонных системах информация, т.е. речь, кодируется в цифровом виде перед передачей по эфиру. Затем закодированная речь декодируется в приемнике. Сначала аналоговый речевой сигнал кодируется в цифровом виде с использованием, к примеру, импульсно-кодовой модуляции (ИКМ). Затем речевыми кодерами и декодерами осуществляется речевое кодирование и декодирование ИКМ речи (или исходной речи). Вследствие возрастания использования радиотелефонных систем доступный для таких систем радиоспектр становится тесным. Для того чтобы обеспечить использование доступного радиоспектра наилучшим возможным образом, радиотелефонные системы используют методы речевого кодирования, которые требуют малого числа разрядов при кодировании речи для сужения требуемой при передаче полосы частот. Постоянно предпринимаются попытки снизить число разрядов, требуемых при речевом кодировании для дальнейшего уменьшения необходимой для передачи речи полосы пропускания.

Известный способ речевого кодирования/декодирования основан на методах линейного кодирования с предсказанием (ЛКП) и использует кодирование возбуждения с анализом через синтез. В использующем такой способ кодере речевой отсчет сначала анализируется для выделения параметров, которые представляют такие характеристики, как информация (ЛКП) о форме сигналов речевого отсчета. Эти параметры используются как входы в синтезирующий фильтр с малой постоянной времени. Синтезирующий фильтр с малой постоянной времени возбуждается сигналами, которые выделены из кодового словаря сигналов. Эти сигналы возбуждения могут быть случайными, например, от словаря стохастических кодов, либо могут быть адаптивными или специально оптимизированными для использования в речевом кодировании. Обычно кодовый словарь содержит две части - фиксированный кодовый словарь и адаптивный кодовый словарь. Выходы возбуждения соответствующих кодовых словарей объединяются, и полное возбуждение поступает в синтезирующий фильтр с малой постоянной времени. Каждый сигнал полного возбуждения фильтруется, и результат сравнивается с исходным речевым отсчетом (закодированным ИКМ) для выделения "ошибки" или разности между синтезированным речевым отсчетом и исходным речевым отсчетом. Полное возбуждение, которое приводит к наименьшей ошибке, выбирается в качестве возбуждения для представления речевого отсчета. Кодово-словарные указатели, или адреса

местоположения соответствующих частичных оптимальных сигналов возбуждения в фиксированном и адаптивном кодовом словаре передаются на приемник вместе с параметрами или коэффициентами ЛКП. Составной кодовый словарь, такой же как в передатчике, находится и в приемнике, и переданные кодово-словарные указатели и параметры используются для генерирования соответствующего сигнала полного возбуждения из кодового словаря приемника. Этот сигнал полного возбуждения подается затем в синтезирующий фильтр с малой постоянной времени, идентичный такому же фильтру в передатчике и имеющий переданные коэффициенты ЛКП в качестве входов. Выход из синтезирующего фильтра с малой постоянной времени представляет собой кадр синтезированной речи, который является тем же самым, что и генерируемый в передатчике способом анализа через синтез. Хотя синтезированная речь объективно точна, она звучит искусственно вследствие природы цифрового кодирования. Кроме того, в синтезированную речь вносятся ослабления, искажения и артефакты из-за эффектов квантования и других аномалий вследствие электронной обработки. Такие артефакты, в частности, происходят при малоразрядном кодировании, поскольку информации для точного воспроизведения исходной речи недостаточно. Поэтому предпринимались попытки улучшить воспринимаемое качество синтезированной речи. Это пытались осуществить путем использования пост-фильтров, которые работают на синтезированных отсчетах для улучшения воспринимаемого качества. Известные пост-фильтры расположены на выходе декодера и обрабатывают сигнал синтезированной речи, чтобы подчеркнуть или ослабить то, что в общем случае рассматривается как наиболее важные частотные области в речи. Важность соответствующих областей речевых частот проанализирована заранее путем использования субъективных тестов на качество результирующего речевого сигнала для человеческого уха. Речь можно разделить на две основные части: спектральную огибающую (формантную структуру) или структуру спектральных гармоник (линейную структуру), и обычно пост-фильтрация подчеркивает одну или другую, либо обе эти части речевого сигнала. Фильтровые коэффициенты пост-фильтра адаптируются в зависимости от характеристики речевого сигнала для согласования звуков речи. Фильтр, подчеркивающий или ослабляющий гармоническую структуру, обычно называется пост-фильтром с большой постоянной времени, или пост-фильтром основного тона, или пост-фильтром длительной задержки, а фильтр, подчеркивающий структуру спектральной огибающей, обычно называется пост-фильтром краткой задержки, или пост-фильтром с малой постоянной времени. Известный, кроме того, способ фильтрации для улучшения воспринимаемого качества синтезированной речи рассматривается в международной патентной заявке WO 91/06091. В заявке WO 91/06091 рассматривается предварительный фильтр основного тона, содержащий фильтр улучшения основного тона, обычно

расположенный в позиции после речевого синтеза или фильтра ЛКП, перемещенный в позицию перед речевым синтезом или фильтром ЛКП, где он фильтрует информацию основного тона, содержащуюся в сигналах возбуждения, входящих в речевой синтез или фильтр ЛКП.

Однако существует все же желание получить синтезированную речь, которая имеет еще лучшее воспринимаемое качество.

Согласно первому аспекту данного изобретения, имеется синтезатор речи ЛКП типа, содержащий пост-процессорное средство для работы на первом сигнале, включающем информацию о периодичности речи, выделенную из источника сигнала возбуждения,

где источник сигнала возбуждения содержит фиксированный кодовый словарь и адаптивный кодовый словарь, и средство получения первого сигнала путем комбинирования первого и второго сигналов частичного возбуждения, происходящих из фиксированного и адаптивного кодовых словарей,

где пост-процессорное средство способно видоизменять содержание информации о периодичности речи первого сигнала в соответствии со вторым сигналом, происходящим из источника сигнала возбуждения, посредством содержания средства регулирования коэффициента усиления для масштабирования второго сигнала в соответствии с первым коэффициентом масштабирования (r), выделяемым из информации основного тона, связанной с первым сигналом, и средства для комбинирования второго сигнала с первым сигналом.

Согласно второму аспекту данного изобретения, имеется способ последующей обработки для улучшения ЛКП-синтезированной речи, включающий в себя этапы выделения первого сигнала, включающего информацию о периодичности речи, из источника сигнала возбуждения, причем источник сигнала возбуждения содержит фиксированный кодовый словарь и адаптивный кодовый словарь, получения первого сигнала путем комбинирования первого и второго сигналов частичного возбуждения, исходящих из фиксированного и адаптивного кодовых словарей, видоизменения содержания информации о периодичности речи первого сигнала в соответствии со вторым сигналом, происходящим из источника сигнала возбуждения путем масштабирования второго сигнала в соответствии с первым коэффициентом масштабирования, выделенным из информации основного тона, связанной с первым сигналом, и комбинирования второго сигнала с первым сигналом.

Преимущество настоящего изобретения состоит в том, что первый сигнал видоизменяется вторым сигналом, возникающим из того же самого источника, что и первый сигнал, тем самым не вводится никаких дополнительных источников искажений или артефактов, таких, как излишние фильтры. Используются лишь сигналы, генерируемые в источнике возбуждения. Относительные вклады сигналов, присущих генератору возбуждения,

в речевом синтезаторе видоизменяются для изменения масштаба синтезируемых сигналов в отсутствие искусственно добавочных сигналов.

Хорошее улучшение речи можно получить, если пост-обработка возбуждения основана на видоизменении относительных вкладов компонент возбуждения, выделенных в генераторе возбуждения самого речевого синтезатора.

Обработка возбуждения путем фильтрации полного возбуждения $e_h(n)$ без рассмотрения или видоизменения относительных вкладов сигналов, присущих генератору возбуждения, т. е. $v(n)$ и $c_v(n)$ обычно не дает наилучшего возможного улучшения. Видоизменение первого сигнала согласно второму сигналу от того же самого источника возбуждения повышает непрерывность формы сигнала в возбуждении и в результирующем синтезированном речевом сигнале, тем самым улучшая воспринимаемое качество.

В предпочтительном выполнении источник возбуждения содержит фиксированный кодовый словарь и адаптивный кодовый словарь, при этом первый сигнал выделяется из комбинации первого и второго частичных сигналов возбуждения, соответственно выбираемых из фиксированного и адаптивного кодовых словарей, что представляет собой чрезвычайно удобный источник возбуждения для речевого синтезатора.

Предпочтительно, имеется усилительный элемент для масштабирования второго сигнала согласно масштабному коэффициенту (r), выделяемому из информации основного тона, связанной с первым сигналом из источника возбуждения, что имеет преимущество, т.к. содержание информации о периодичности речи первого сигнала видоизменяется, что имеет больший эффект на воспринимаемое качество речи, чем иные видоизменения.

Соответственно масштабный коэффициент (r) выделяется из масштабного коэффициента (b) адаптивного кодового словаря, и масштабный коэффициент (r) выделяется в соответствии со следующим уравнением:

$$\begin{aligned} b < \Pi_{\text{нижн.}} \text{, то } r = 0,0, \\ \Pi_{\text{нижн.}} \leq b < \Pi_2, \text{ то } r = a_{\text{улучш}} f_1(b), \\ \text{если } \Pi_2 \leq b < \Pi_3, \text{ то } r = a_{\text{улучш}} f_2(b), \\ \Pi_{N-1} \leq b < \Pi_{\text{верх.}}, \text{ то } r = a_{\text{улучш}} f_{N-1}(b), \\ b > \Pi_{\text{верх.}}, \text{ то } r = a_{\text{улучш}} f_N(b), \end{aligned}$$

где Π представляет пороговые значения, b является коэффициентом усиления адаптивного кодового словаря, r представляет собой масштабный коэффициент пост-процессорного средства, $a_{\text{улучш}}$ является линейным множителем, $f(b)$ есть функция от усиления b .

В конкретном выполнении масштабный коэффициент (r) выделяется согласно

$$\begin{aligned} b < \Pi_{\text{нижн.}} \text{, то } r = 0,0, \\ \text{если } \Pi_{\text{нижн.}} \leq b \leq \Pi_{\text{верх.}}, \text{ то } r = a_{\text{улучш}} b^2, \\ b > \Pi_{\text{верх.}}, \text{ то } r = a_{\text{улучш}} b, \end{aligned}$$

где $a_{\text{улучш}}$ представляет собой постоянную, которая управляет интенсивностью операции улучшения, b является усилением адаптивного кодового словаря, Π есть пороговые значения, a r

является пост-процессорным масштабным коэффициентом, который использует понимание того, что улучшение речи наиболее эффективно для огласованной речи, где b обычно имеет высокое значение, тогда как для неогласованных звуков, где b имеет низкое значение, требуется не настолько сильное улучшение.

Второй сигнал может возникать из адаптивного кодового словаря и может также быть практически тем же самым, что и второй частичный сигнал возбуждения. Альтернативно, второй сигнал может возникать из фиксированного кодового словаря и может быть также практически тем же самым, что и первый частичный сигнал возбуждения.

Для второго сигнала, возникающего из фиксированного кодового словаря, средство регулирования усиления приспособлено масштабировать второй сигнал согласно второму коэффициенту (r'), где

$$r' = - \frac{gb}{(p+b)},$$

где g является масштабным коэффициентом фиксированного кодового словаря, b есть масштабный коэффициент адаптивного кодового словаря.

Первый сигнал может быть первым сигналом возбуждения, пригодным для введения в речевой синтезирующий фильтр, а второй сигнал может быть вторым сигналом возбуждения, пригодным для введения в речевой синтезирующий фильтр. Второй сигнал возбуждения может быть практически тем же самым, что и второй частичный сигнал возбуждения.

В некоторых случаях первый сигнал может быть выходом первого синтезированного речевого сигнала из первого речевого синтезирующего фильтра, выделяемым из первого сигнала возбуждения, а второй сигнал может быть выходом из второго речевого синтезирующего фильтра, выделяемым из второго сигнала возбуждения. Преимущество этого в том, что улучшение речи осуществляется над действительно синтезированной речью, и тем самым имеется меньше электронных компонент, влияющих на внесение искажений в сигнал перед тем, как он воспроизводится в звуке.

Выгодно, чтобы предусматривалось адаптивное средство управления энергией, приспособленное для масштабирования видеоизмененного первого сигнала согласно следующему соотношению:

$$k = \frac{\sum_{n=0}^{N-1} e x^2(n)}{\sum_{n=0}^{N-1} e w'^2(n)}$$

где N есть соответственно выбранный период адаптации, $e x(n)$ представляет собой первый сигнал, $e w'(n)$ есть видеоизмененный первый сигнал, k является масштабным коэффициентом энергии, который нормирует результирующий улучшенный сигнал к мощности входа в речевой синтезатор.

В третьем аспекте согласно этому изобретению предлагаются радиоустройство, содержащее:

высокочастотное средство для приема радиосигнала и восстановления закодированной информации, содержащейся в этом радиосигнале, и синтезатор в соответствии с любым из пп.1-14.

В четвертом аспекте изобретения имеется синтезатор речи ЛКП-типа, включающий в себя:

адаптивный кодовый словарь и фиксированный кодовый словарь для генерирования первого и второго сигналов частичного возбуждения, соответственно,

средство масштабирования для масштабирования первого и второго сигналов частично возбуждения посредством коэффициентов масштабирования,

полученных из адаптивного и фиксированного кодовых словарей, соответственно, видоизменяющее средство для видоизменения первого сигнала возбуждения в соответствии с еще одним коэффициентом масштабирования, причем коэффициент масштабирования является функцией

информации основного тона, связанной с первым сигналом возбуждения, и средство комбинирования второго сигнала частичного возбуждения с видоизмененным первым сигналом частичного возбуждения.

В пятом аспекте имеется синтезатор речи ЛКП-типа, включающий в себя: адаптивный кодовый словарь и фиксированный кодовый словарь для генерирования первого и второго сигналов частичного возбуждения, соответственно,

средство масштабирования для масштабирования первого и второго сигналов частичного возбуждения посредством коэффициентов масштабирования,

полученных из адаптивного и фиксированного кодовых словарей, соответственно, видоизменяющее средство для видоизменения второго сигнала возбуждения в соответствии с еще одним коэффициентом масштабирования, причем коэффициент масштабирования является функцией

информации основного тона, связанной с первым сигналом возбуждения, и

средство комбинирования видоизмененного второго сигнала частичного возбуждения с первым сигналом частичного возбуждения.

Четвертый и пятый аспекты изобретения выгодно интегрируют масштабирование сигналов возбуждения в самом генераторе возбуждения.

Рассмотрим теперь выполнения согласно изобретению посредством только примеров и со ссылками на сопровождающие чертежи.

Фиг. 1 показывает схему известного кодера кодовых возбуждений с линейным предсказанием (КВЛП).

Фиг.2 показывает схему известного декодера КВЛП.

Фиг.3 показывает схему декодера КВЛП согласно первому варианту выполнения изобретения.

Фиг.4 показывает второй вариант выполнения согласно изобретению.

Фиг.5 показывает третий вариант выполнения согласно изобретению.

Фиг.6 показывает четвертый вариант выполнения согласно изобретению.

Фиг.7 показывает пятый вариант выполнения согласно изобретению.

На фиг.1 показан известный кодер 100

КВЛП. Исходные речевые сигналы входят в кодер 102, и коэффициенты T , b долговременного предсказания (ДВП) определяются с использованием адаптивного кодового словаря 104. Эти коэффициенты ДВП определяются для сегментов речи, обычно содержащих 40 отсчетов, и имеют длину 5 мс. Эти коэффициенты ДВП относятся к периодическим характеристикам исходной речи. Это включает в себя любую периодичность в исходной речи, а не только ту периодичность, которая соответствует основному тону исходной речи вследствие колебаний голосовых связок человека, произносящего исходную речь.

Долговременное предсказание выполняется с использованием адаптивного кодового словаря 104 и усилительного элемента 114, который содержит часть генератора 126 сигнала $(ex(n))$ возбуждения, показанного пунктиром на фиг.1. Предыдущие сигналы $ex(n)$ возбуждения запоминаются в адаптивном кодовом словаре 104 посредством петли 122 обратной связи. Во время процесса ДВП адаптивный кодовый словарь просматривается путем изменения адреса T , известного как задержка или запаздывание, указывающего предыдущие сигналы $ex(n)$ возбуждения. Эти сигналы последовательно выводятся и усиливаются в усилительном элементе 114 с масштабным коэффициентом b для образования сигналов $v(n)$ перед добавлением в элементе 118 к сигналу $c_i(n)$ возбуждения, выделенному из фиксированного кодового словаря 112 и умноженному на коэффициент g в усилительном элементе 116. Коэффициенты линейного предсказания (ЛП) для речевого отсчета вычисляются в элементе 106. Коэффициенты ЛП квантуются затем в элементе 108. Квантованные коэффициенты ЛП доступны затем для передачи по эфиру и для введения в фильтр 110 с малой постоянной времени. Коэффициенты ЛП $(r(i))$, $i=1, \dots, m$, где m является порядком предсказания вычисляются для сегментов речи, содержащих 160 отсчетов на 20 мс. Вся дальнейшая обработка обычно выполняется в сегментах из 40 отсчетов, т.е. на длине кадра возбуждения в 5 мс. Коэффициенты ЛП относятся к спектральной огибающей исходного речевого сигнала.

Генератор 126 возбуждения фактически содержит составной кодовый словарь 104, 112, содержащий набор кодов для возбуждения синтезирующего фильтра 110 с малой постоянной времени. Эти коды содержат последовательности амплитуд напряжения, каждая из которых соответствует речевому отсчету в речевом кадре.

Каждый сигнал $ex(n)$ полного возбуждения является входом для синтезирующего фильтра 110 ЛКП или с малой постоянной времени для образования синтезированного речевого отсчета $s(n)$. Этот синтезированный речевой отсчет $s(n)$ является входом для отрицательного входа сумматора 120, положительным входом для которого является исходный речевой отсчет. Сумматор 120 выдает разность между исходным речевым отсчетом и синтезированным речевым отсчетом, причем эта разность известна как объективная ошибка. Эта объективная ошибка вводится в элемент 124 выбора наилучшего возбуждения, который

выбирает полное возбуждение $ex(n)$, проявляющееся в синтезированном речевом кадре $s(n)$ с наименьшей объективной ошибкой. В процессе этого выбора объективная ошибка далее обычно взвешивается для подчеркивания тех спектральных областей речевого сигнала, которые важны для человеческого восприятия. Затем соответствующие параметры адаптивного и фиксированного кодовых словарей (усиление b и задержка T , а также усиление g и указатель i), дающие сигнал $ex(n)$ наилучшего возбуждения, передаются вместе с коэффициентами $r(i)$ фильтра ЛКП на приемник для использования в синтезировании речевого кадра для восстановления исходного речевого сигнала.

На фиг. 2 показан декодер, пригодный для декодирования речевых параметров, генерируемых кодером, описанным со ссылкой на фиг.1. Высокочастотный (ВЧ) блок 201 принимает кодированный речевой сигнал через антенну 212. Принятый высокочастотный сигнал преобразуется с понижением на частоту модулирующих сигналов и демодулируется в ВЧ блоке 201 для восстановления речевой информации. В общем случае, кодированная речь дополнительно кодируется перед передачей, чтобы включать в себя канальное кодирование и кодирование с исправлением ошибок. Это канальное кодирование и кодирование с исправлением ошибок должно декодироваться в приемнике перед тем, как можно обратиться к речевому кодированию или выделить его. Параметры речевого кодирования выделяются параметрическим декодером 202. Параметры речевого кодирования в речевом кодировании с линейным предсказанием представляют собой набор коэффициентов $r(i)$ синтезирующего фильтра ЛКП ($i=1, \dots, m$, где m - порядок предсказания), указатель i фиксированного кодового словаря и усиление g . Выделяются также такие параметры речевого кодирования адаптивного кодового словаря, как задержка T и усиление b .

Речевой декодер 200 использует вышеупомянутые параметры речевого кодирования для получения от генератора 211 возбуждения сигнала $ex(n)$ возбуждения для введения в синтезирующий фильтр 208 ЛКП, который выдает на своем выходе сигнал $s(n)$ синтезированного речевого кадра в качестве отклика на сигнал $ex(n)$ возбуждения. Сигнал $s(n)$ синтезированного речевого кадра обрабатывается далее в блоке 209 звуковой обработки и выдается в звуковом виде через соответствующий звуковой преобразователь 210.

В обычных речевых декодерах с линейным предсказанием сигнал $ex(n)$ возбуждения для синтезирующего фильтра 208 ЛКП образуется в генераторе 211 возбуждения, содержащем фиксированный кодовый словарь 203, генерирующий последовательность $c_i(n)$ возбуждения, и адаптивный кодовый словарь 204. Положение кодово-словарной последовательности $ex(n)$ возбуждения в соответствующих кодовых словарях 203, 204 указывается параметром i речевого кодирования и задержкой T . Последовательность $c_i(n)$ возбуждения фиксированного кодового словаря, частично используемая для образования сигнала $ex(n)$

возбуждения, берется из фиксированного кодового словаря 203 возбуждения из положения, указанного указателем i , и затем соответственно масштабируется переданным коэффициентом g усиления в масштабирующем блоке 205. Аналогично, последовательность $v(n)$ возбуждения адаптивного кодового словаря, также частично используемая для образования сигнала $ex(n)$ возбуждения, берется из адаптивного кодового словаря 204 из положения, указанного задержкой T , с использованием логики выбора, присущей адаптивному кодовому словарю, а затем соответственно масштабируется переданным коэффициентом b усиления в масштабирующем блоке 206.

Адаптивный кодовый словарь 204 работает на последовательности $c_i(n)$ возбуждения фиксированного кодового словаря путем добавления компоненты $v(n)$ частичного возбуждения к последовательности $g c_i(n)$ возбуждения кодового словаря. Вторая компонента выделяется из прошлых сигналов возбуждения с помощью уже описанного со ссылкой на фиг.1 способа и выбирается из адаптивного кодового словаря 204 с использованием логики выбора, соответственно включенной в адаптивный кодовый словарь. Компонента $v(n)$ соответственно масштабируется в масштабирующем блоке 206 переданным усилием b адаптивного кодового словаря, а затем добавляется к $g c_i(n)$ в сумматоре 207 для образования сигнала $ex(n)$ полного возбуждения, где

$$ex(n) = g c_i(n) + b v(n). \quad (1)$$

Затем адаптивный кодовый словарь 204 обновляется за счет использования сигнала $ex(n)$ полного возбуждения.

Положение второй компоненты $v(n)$ частичного возбуждения в адаптивном кодовом словаре 204 указывается параметром T речевого кодирования. Адаптивная компонента возбуждения выбирается из адаптивного кодового словаря с использованием параметра T речевого кодирования и логики выбора, включенной в адаптивный кодовый словарь.

Декодер 300 речевого синтеза ЛКП согласно изобретению показан на фиг.3. Действие речевого синтеза согласно фиг.3 то же самое, что и для фиг.2, за исключением того, что сигнал $ex(n)$ полного возбуждения перед тем, как быть использованным в качестве возбуждения для синтезирующего фильтра 208 ЛКП, обрабатывается в пост-процессорном блоке 317 (блоке пост-обработки). Действие схемных элементов 201-212 на фиг.3 такое же, как у элементов с теми же позициями на фиг.2.

Согласно аспекту изобретения, в речевом декодере 300 для полного возбуждения $ex(n)$ используется пост-процессорный блок 317. Этот пост-процессорный блок 317 содержит сумматор 313 для добавления третьей компоненты к полному возбуждению $ex(n)$. Затем усилительный блок 315 соответственно масштабирует результирующий сигнал $ew(n)$ для образования сигнала $ew'(n)$, который используется потом для возбуждения синтезирующего фильтра 208 ЛКП, чтобы получить синтезированный речевой сигнал $S_{ew}(n)$. Речевое синтезирование

согласно изобретению улучшает воспринимаемое качество по сравнению с речевым сигналом $s(n)$, синтезированным известным декодером речевого синтеза, показанным на фиг.2.

Пост-процессорный блок 317 имеет вход полного возбуждения $ex(n)$ и выдает полное возбуждение $ew(n)$ улучшенного восприятия. Пост-процессорный блок 317 имеет также усиление b адаптивного кодового словаря и немасштабированную компоненту $v(n)$ частичного возбуждения, которая берется из адаптивного кодового словаря 204 в положении, указанном параметрами речевого кодирования как дополнительными входами. Компонента $v(n)$ частичного возбуждения практически та же самая компонента, которая используется в генераторе 211 возбуждения для образования второй компоненты $b v(n)$ возбуждения, которая добавляется к масштабированному возбуждению $g c_i(n)$ кодового словаря для образования полного возбуждения $ex(n)$. При использовании последовательности возбуждения, которая выделяется из адаптивного кодового словаря 204, никакие источники артефактов не добавляются к электронике речевой обработки, как в случае с известными методами пост- или предварительной фильтрации, которые используют лишние фильтры. Блок 317 пост-обработки возбуждения содержит также масштабирующий блок 314, который масштабирует компоненту $v(n)$ частичного возбуждения масштабным коэффициентом r , и масштабированная компонента $pv(n)$ добавляется сумматором 313 к компоненте $ex(n)$ полного возбуждения. Выход сумматора 313 представляет собой промежуточный сигнал $ew'(n)$ полного возбуждения. Он имеет вид

$$ew'(n) = g c_i(n) + b v(n) + p v(n) = g c_i(n) + (b + p) v(n). \quad (2)$$

Масштабный коэффициент r для масштабирующего блока 314 определяется в блоке 312 регулировки усиления улучшенного восприятия с использованием усиления b адаптивного кодового словаря. Масштабный коэффициент r перемасштабирует вклад двух компонент возбуждения из фиксированного и адаптивного кодовых словарей, соответственно $c_i(n)$ и $v(n)$. Масштабный коэффициент r регулируется так, что во время отсчетов синтезированного речевого кадра, которые имеют высокое значение усиления b адаптивного кодового словаря, этот масштабный коэффициент r увеличивается, в то время речи, которая имеет низкое значение усиления b адаптивного кодового словаря, масштабный коэффициент r снижается. Кроме того, когда b меньше, чем пороговое значение ($b < \Pi_{\text{ниж}}$), масштабный коэффициент r устанавливается на нуль. Блок 312 регулировки усиления улучшенного восприятия работает в соответствии с приведенным ниже уравнением (3).

$$b < \Pi_{\text{ниж}}: r = 0,0,$$

если $\Pi_{\text{ниж}} \leq b \leq \Pi_{\text{верх}}$, $r = a_{\text{улучш}} b^2$,

$$b > \Pi_{\text{верх}}, r = a_{\text{улучш}} b \quad (3),$$

где $a_{\text{улучш}}$ является постоянной, которая управляет интенсивностью операции улучшения. Заявитель обнаружил, что хорошим значением для $a_{\text{улучш}}$ является

0,25, а хорошие значения для $\Pi_{\text{нижн}}$ и $\Pi_{\text{верх}}$ составляют, соответственно, 0,5 и 1,0.

Уравнение (3) может иметь более общий вид, и обобщенная формулировка функции улучшения дана ниже в уравнении (4). В общем случае может быть больше, чем два порога для улучшенного усиления b . Кроме того, усиление можно определить как более общую функцию от b .

$$b < \Pi_{\text{нижн}}, p = 0,0,$$

$$\Pi_{\text{нижн}} \leq b < \Pi_2, p = a_{\text{улучш1}} f_1(b),$$

$$\Pi_2 \leq b < \Pi_3, p = a_{\text{улучш2}} f_2(b),$$

если

$$\Pi_{N-1} \leq b \leq \Pi_{\text{верх}}, p = a_{\text{улучшN-1}} f_{N-1}(b),$$

$$b > \Pi_{\text{верх}}, p = a_{\text{улучшN}} f_N(b).$$

В описанном ранее предпочтительном выполнении $N = 2$, $\Pi_{\text{нижн}} = 0,5$, $\Pi_2 = 1,0$, $\Pi_3 = \infty$, $a_{\text{улучш1}} = 0,25$, $a_{\text{улучш2}} = 0,25$, $f_1(b) = b^2$, $f_2(b) = b$.

Пороговые значения (Π), улучшенные значения ($a_{\text{улучш}}$) и функции ($f(b)$) усиления получены эмпирически. Поскольку единственную естественную меру качества воспринимаемой речи можно получить людьми, прослушивающими речь и дающими их субъективные мнения по качеству этой речи, значения, использованные в уравнениях (3) и (4), определены экспериментально. Пробовались различные значения для улучшенных порогов и функций усиления, и выбирались их результаты в наилучшем звучании речи. Заявитель использовал понимание того, что улучшение в качестве речи при использовании этого способа особенно эффективно для огласованной речи, где b обычно имеет высокое значение, тогда как для менее огласованных звуков, которые имеют более низкое значение b , не требуется такого сильного улучшения. Таким образом, значение p управляется так, что для огласованных звуков, где искажения наиболее слышны, эффект силен, а для неогласованных звуков эффект слабее или не используется вовсе. Таким образом, как общее правило, функции (f_n) должны выбираться так, чтобы больший эффект был для более высоких значений b , чем для более низких значений b . Это увеличивает разность между компонентами основного тона речи и иными компонентами.

В предпочтительном варианте выполнения, работающем в соответствии с уравнением (3), функции, меняющиеся от значения b усиления, имеют квадратичную зависимость от средних по диапазону значений b и линейную зависимость от высоких по диапазону значений b . Нынешнее понимание заявителя состоит в том, что это дает хорошее качество речи, т.к. для высоких значений b , т.е. высоко огласованной речи, эффект больше, а для низких значений b эффект меньше. Это имеет место потому, что b лежит обычно в диапазоне $-1 < b < 1$ поэтому $b^2 < b$.

Чтобы обеспечить единое усиление мощности между входным сигналом $ex(n)$ и выходным сигналом $ew(n)$ блока 317 пост-обработки возбуждения, масштабный коэффициент вычисляется и используется в масштабирующем блоке 315 для масштабирования промежуточного сигнала $ew(n)$ возбуждения, чтобы получить пост-процессорный сигнал $ew(n)$

возбуждения. Масштабный коэффициент k задан как

$$k = \sqrt{\frac{\sum_{n=0}^{N-1} ex^2(n)}{\sum_{n=0}^{N-1} ew^2(n)}} \quad (5)$$

где N представляет собой соответственно выбранный период адаптации. Обычно N устанавливается равным длине кадра возбуждения речевого кодека ЛКП.

В адаптивном кодовом словаре кодера для значений T , которые меньше, чем длина кадра или длина возбуждения, часть последовательности возбуждения неизвестна. Для этих неизвестных частей в адаптивном кодовом словаре генерируется на месте заменяющая последовательность путем использования соответствующей логики выбора. Из уровня техники известно несколько методов в адаптивном кодовом словаре для генерирования этой заменяющей последовательности. Обычно копия части известного возбуждения копируется туда, где расположена неизвестная часть, благодаря чему создается законченная последовательность возбуждения.

Скопированная часть может каким-либо образом приспособливаться для улучшения качества результата генерации речевого сигнала. При выполнении такой копии значение задержки T не используется, т.к. оно указывало бы неизвестную часть. Вместо этого используется конкретная логика выбора, дающая в результате видоизмененное значение для T (например, с использованием T , умноженного на целочисленный коэффициент, так что оно всегда указывает на известную часть сигнала). Поскольку декодер синхронизируется с кодером, сходные модификации осуществляются и в адаптивном кодовом словаре декодера. За счет использования в адаптивном кодовом словаре логики выбора для генерирования заменяющей последовательности, этот адаптивный кодовый словарь способен адаптироваться к голосам с высоким основным тоном, таким, как женские и детские голоса, что приводит в результате к эффективному генерированию возбуждения и улучшенному качеству речи для этих голосов.

Для получения хорошего улучшения восприятия, в улучшенной пост-обработке принимаются во внимание все видоизменения, присущие адаптивному кодовому словарю, например, для значений T меньше, чем длина кадра. Согласно изобретению, это достигается использованием последовательности $v(n)$ частичного возбуждения из адаптивного кодового словаря и пересыпыванием компонент возбуждения, присущих генератору возбуждения речевого синтезатора.

Вкратце, способ улучшает воспринимаемое качество синтезированной речи и снижает звуковые артефакты за счет адаптивного масштабирования, в соответствии с уравнениями (2), (3), (4) и (5), вклада компонент частичного возбуждения, взятых из кодового словаря 203 и из адаптивного кодового словаря 204.

Фиг.4 показывает второй вариант

выполнения согласно изобретению, в котором блок 417 пост-обработки возбуждения расположен после синтезирующего фильтра 208 ЛКП, как представлено. В этом выполнении дополнительный синтезирующий фильтр 408 ЛКП требуется для третьей компоненты возбуждения, которая выделяется из адаптивного кодового словаря 204. На фиг.4 элементы, которые имеют те же самые позиции, что и на фиг.2 и 3, также имеют те же самые функции.

Во втором варианте выполнения, показанном на фиг.4, синтезирующий фильтр ЛКП улучшает восприятие пост-процессором 417. Сигнал $ex(n)$ полного восприятия, выделенный из кодового словаря 203 и адаптивного кодового словаря 204, вводится в синтезирующий фильтр 208 ЛКП и обрабатывается обычным образом в соответствии с коэффициентами $r(i)$. Дополнительная или третья компонента $v(n)$ частичного возбуждения, выделенная из адаптивного кодового словаря 204 способом, описанным в отношении фиг.3, вводится без масштабирования во второй синтезирующий фильтр 408 ЛКП и обрабатывается в соответствии с коэффициентами $r(i)$. Выходы $s(n)$ и $sv(n)$ соответствующих фильтров 208, 408 ЛКП вводятся в пост-процессор 417 и складываются в сумматоре 413. Перед введением в сумматор 413 сигнал $sv(n)$ масштабируется масштабным коэффициентом r . Как описано для фиг.3, значения для масштабного коэффициента обработки или усиления r можно получить эмпирически. Вдобавок, третья компонента частичного возбуждения может быть выделена из фиксированного кодового словаря 203, а масштабированный речевой сигнал $r'sv(n)$ вычитается из речевого сигнала $s(n)$.

Результирующий выход $s_w(n)$ улучшенного восприятия вводится затем в блок 209 звуковой обработки.

Можно выполнить дальнейшую модификацию улучшенной системы путем перемещения масштабирующего блока 414 на фиг.4 в положение перед синтезирующим фильтром 408 ЛКП. Расположение пост-процессора 417 после синтезирующих фильтров 208, 408 ЛКП или с малой постоянной времени может обеспечить лучшее управление подчеркиванием речевого сигнала, поскольку оно выполняется прямо на речевом сигнале, а не на сигнале возбуждения. Тем самым, вероятно, будет меньше искажений.

Улучшения можно достичь и такой модификацией вариантов, описанных со ссылками на фиг. 3 и 4, что дополнительная (третья) компонента возбуждения выделяется из фиксированного кодового словаря 203 вместо адаптивного кодового словаря 204. Затем следует использовать отрицательный масштабный коэффициент вместо исходного положительного коэффициента r усиления, чтобы понизить усиление для последовательности $c_i(n)$ возбуждения из фиксированного кодового словаря. Это приводит к такому же видоизменению относительных вкладов сигналов $c_i(n)$ и $v(n)$ частичного возбуждения в речевой синтез, как и достигаемое вариантами по фиг.3 и 4.

Фиг.5 показывает вариант выполнения

согласно изобретению, в котором можно достичь того же результата, что и при использовании масштабного коэффициента r и дополнительной компоненты возбуждения из адаптивного кодового словаря. В этом варианте выполнения последовательность $c_i(n)$ возбуждения фиксированного кодового словаря вводится в масштабирующий блок 314, который работает в соответствии с масштабным коэффициентом r' , выводимым из второй регулировки 512 усиления улучшенного восприятия. Масштабированное возбуждение $r'c_i(n)$ фиксированного кодового словаря, выводимое из масштабирующего блока 314, вводится в сумматор 313, где оно добавляется к последовательности $ex(n)$ полного возбуждения, содержащей компоненты $c_i(n)$ и $v(n)$ из фиксированного кодового словаря 203 и адаптивного кодового словаря 204, соответственно.

При увеличении усиления для сигнала $v(n)$ последовательности возбуждения из адаптивного кодового словаря 204 полное возбуждение (перед адаптивной регулировкой 316 энергии) задается уравнением (2), а именно:

$$ew'(n) = gc_i(n) + (b + p)v(n). \quad (2)$$

При понижении усиления для последовательности $c_i(n)$ возбуждения из фиксированного кодового словаря 203 полное возбуждение (перед адаптивной регулировкой 316 энергии) задается как

$$ew'(n) = (g + p')c_i(n) + bv(n), \quad (6)$$

где p' представляет собой масштабный коэффициент, выделенный из второй регулировки 512 усиления улучшенного восприятия, показанной на фиг.5. Взяв уравнение (2) и переписав его в виде, аналогичном уравнению (6), получим:

$$\begin{aligned} ew'(n) &= gc_i(n) + (b + p)v(n) = \\ &= \frac{p+b}{b} \left[\left(\frac{gb}{p+b} \right) c_i(n) + bv(n) \right] = \\ &= \frac{p+b}{b} \left[\left(g - \frac{gp}{p+b} \right) c_i(n) + bv(n) \right]. \end{aligned}$$

Таким образом, выбирая

$$p' = - \frac{gp}{(p+b)},$$

в варианте по фиг.5 получается такое же улучшение, как и достигнутое в варианте по фиг. 3. Когда промежуточный сигнал $ew'(n)$ масштабируется адаптивной регулировкой 316 энергии до такой же величины энергии, что и $ex(n)$, оба варианта выполнения, на фиг.3 и фиг.5, дают один и тот же сигнал $ew(n)$ полного возбуждения.

Вторая регулировка 512 усиления улучшенного восприятия может поэтому использовать ту же самую обработку, которая используется в отношении вариантов по фиг. 3 и 4, для генерирования " r' ", а затем использовать уравнение (8) для получения r .

Промежуточный сигнал $ew'(n)$ полного возбуждения, выводимый из сумматора 313, масштабируется в масштабном блоке 315 под управлением адаптивной регулировки 316 энергии так же, как описано выше в отношении первого и второго вариантов выполнении.

На фиг.4 синтезированная речь ЛКП может улучшаться по восприятию пост-процессором 417 с помощью синтезированной речи, выделенной из

дополнительных сигналов возбуждения из фиксированного кодового словаря.

Пунктир 420 на фиг.4 показывает вариант выполнения, в котором сигналы $c_i(n)$ возбуждения фиксированного кодового словаря соединяются с синтезирующим фильтром 408 ЛКП. Выход синтезирующего фильтра 408 ЛКП ($sc_i(n)$) затем масштабируется в блоке 414 в соответствии с масштабным коэффициентом r' , выделенным из регулировки 512 усиления улучшенного восприятия, и добавляется к синтезированному сигналу $s(n)$ в сумматоре 413 для получения промежуточного синтезированного сигнала $s'_w(n)$. После нормировки в масштабирующем блоке 415 результирующий синтезированный сигнал $s_w(n)$ подается на блок 209 звуковой обработки.

Предыдущие варианты выполнения содержат добавление компоненты, выделенной из адаптивного кодового словаря 204 или фиксированного кодового словаря 203 к возбуждению $ex(n)$ или синтезированному $s(n)$ для образования промежуточного возбуждения $ew'(n)$ или синтезированного сигнала $s'_w(n)$.

Можно обойтись и без пост-обработки, а сигналы $c_i(n)$ и $v(n)$ возбуждения адаптивного кодового словаря или фиксированного кодового словаря могут масштабироваться и объединяться непосредственно. Тем самым устраняется добавление компонент к немасштабированным объединенным сигналам фиксированного и адаптивного кодовых словарей.

Фиг. 6 показывает вариант выполнения согласно аспекту изобретения с сигналами $v(n)$ возбуждения адаптивного кодового словаря, масштабированными и затем объединенными с сигналами $c_i(n)$ возбуждения фиксированного кодового словаря для непосредственного образования промежуточного сигнала $ew'(n)$.

Регулировка 612 усиления улучшенного восприятия выдает параметр "a" для управления масштабирующим блоком 614. Масштабирующий блок 614 работает по сигналу $v(n)$ возбуждения адаптивного кодового словаря для растягивания или усиления сигнала $v(n)$ возбуждения с помощью коэффициента b усиления, используемого для получения нормального возбуждения. Нормальное возбуждение $ex(n)$ образуется и соединяется с адаптивным кодовым словарем 204 и адаптивной регулировкой 316 энергии. Сумматор 613 объединяет растянутый сигнал $av(n)$ возбуждения и возбуждение с $c_i(n)$ фиксированного кодового словаря, чтобы получить промежуточный сигнал:

$$ew'(n) = g c_i(n) + av(n). \quad (9)$$

Если $a = b+r$, то можно достичь той обработки, которая задается уравнением (2).

Фиг.7 показывает вариант выполнения, работающий аналогично тому, который показан на фиг. 6, но осуществляющий сжатие или ослабление сигнала возбуждения $c_i(n)$ фиксированного кодового словаря. Для этого варианта промежуточный сигнал $ew'(n)$ возбуждения задается:

$$ew'(n) = (g + r') c_i(n) + bv(n) = a'c_i(n) + bv(n), \quad (10)$$

$$\text{где } a' = g - \frac{gr}{r+b} = \frac{gb}{r+b}. \quad (11)$$

Регулировка 712 усиления улучшенного восприятия выводит управляющий сигнал a' в соответствии с уравнением (11), чтобы получить тот же результат, что и полученный с помощью уравнения (6) в соответствии с уравнением (8). Сжатый сигнал $a'c_i(n)$ объединяется с сигналом $v(n)$ возбуждения адаптивного кодового словаря в сумматоре 713 для образования промежуточного сигнала $ew'(n)$ возбуждения. Остальная обработка выполняется, как описано выше, для нормирования сигнала возбуждения и образованного синтезированного сигнала $s_{ew}(n)$.

Варианты, описанные со ссылками на фиг.6 и 7, осуществляют масштабирование сигналов возбуждения в генераторе возбуждения и прямо из кодовых словарей.

Определение масштабирующего коэффициента "r" для вариантов, описанных со ссылками на фиг.5, 6 и 7, можно производить согласно уравнениям (3) или (4), описанным выше.

Можно использовать различные способы управления уровнем улучшения (a улучш). В дополнение к усилению b адаптивного кодового словаря, величина улучшения может быть функцией от значения T отставания или задержки для адаптивного кодового словаря 204. К примеру, пост-обработка может включаться (или подчеркиваться) при работе в диапазоне высоких основных тонов, либо когда параметр T адаптивного кодового словаря короче, чем длина блока возбуждения (фактического диапазона отставания). В результате будут подвергаться наибольшей пост-обработке женские и детские голоса, для которых изобретение наиболее выигрышно.

Управление пост-обработкой может также основываться на решениях об огласованной/неогласованной речи. К примеру, улучшение может быть сильнее для огласованной речи и оно может полностью выключаться, когда речь классифицируется как неогласованная. Это можно выделить из значения b усиления адаптивного кодового словаря, которое само по себе является простой мерой огласованной/неогласованной речи, - иначе говоря, чем выше b , тем более огласованная речь присутствует в исходном речевом сигнале.

Варианты согласно настоящему изобретению можно видоизменять, так что третья последовательность частичного возбуждения не будет той самой последовательностью частичного возбуждения, выделенной из адаптивного кодового словаря или фиксированного кодового словаря в соответствии с обычным синтезом речи, но может выбираться с помощью логики выбора, обычно включенной в соответствующие кодовые словари для выбора другой третьей последовательности частичного возбуждения. Эта третья последовательность частичного возбуждения может выбираться так, чтобы быть непосредственно перед этим использованной последовательностью возбуждения, либо всегда быть одной и той же последовательностью возбуждения, запомненной в фиксированном кодовом

словаре. Это приведет к уменьшению разности между речевыми кадрами и, тем самым, к улучшению непрерывности речи. Факультативно, b или/и T можно пересчитать в декодере из синтезированной речи и использовать для выделения третьей последовательности частичного возбуждения. Кроме того, к последовательности $ex(n)$ полного возбуждения или к речевому сигналу $s(n)$ можно добавить фиксированное усиление r или/и фиксированную последовательность возбуждения, либо при необходимости вычесть их из последовательности $ex(n)$ полного возбуждения или из речевого сигнала $s(n)$ в зависимости от положения пост-процессора.

Ввиду предыдущего описания специалисту будет ясно, что можно сделать различные видоизменения в объеме изобретения. К примеру, в кодере можно использовать кодирование с переменной скоростью кадров, быстрый поиск кодового словаря, реверсирование порядка предсказания основного тона и ЛКП. В дополнение к этому, последующая обработка в соответствии с данным изобретением могла бы тоже включаться в кодер, а не только в декодер. Кроме того, аспекты соответствующих вариантов выполнении, описанные со ссылкой на чертежи, можно объединить для получения дальнейших вариантов согласно изобретению.

Формула изобретения:

1. Синтезатор для речевого синтеза, действующий на основании коэффициентов линейного предсказания и содержащий постпроцессорное средство для обработки первого сигнала ($ex(n)$), включающего в себя информацию о периодичности речи, выделенного из источника сигнала возбуждения, содержащего фиксированный кодовый словарь и адаптивный кодовый словарь, и средство для получения первого сигнала путем комбинирования первого и второго сигналов частичного возбуждения, исходящих соответственно из фиксированного и адаптивного кодовых словарей, причем постпроцессорное средство выполнено с возможностью видоизменять содержание информации о периодичности речи первого сигнала в соответствии с вторым сигналом, выделяемым из источника сигнала возбуждения, благодаря наличию средства регулировки усиления для масштабирования второго сигнала в соответствии с первым коэффициентом (r) масштабирования, выделенного из информации основного тона, связанной с первым сигналом, и средства для комбинирования второго сигнала с первым сигналом.

2. Синтезатор по п. 1, отличающийся тем, что он выполнен с возможностью выделения первого коэффициента (r) масштабирования из коэффициента (b) масштабирования адаптивного кодового словаря.

3. Синтезатор по п. 2, отличающийся тем, что он выполнен с обеспечением выделения первого коэффициента (r) масштабирования из следующего соотношения:

$$b < \Pi_{нижн} r - 0,0$$

$$\Pi_{нижн} \leq b < \Pi_2 r - a_{улучш1} f_1(b)$$

$$\Pi_2 \leq b < \Pi_3 r - a_{улучш2} f_2(b)$$

если

$$\Pi_{нижн} \leq b \leq \Pi_{верх} r - a_{улучшN-1} f_{N-1}(b)$$

$$b > \Pi_{верх} r - a_{улучшN} f_N(b)$$

где Π - пороговые значения;

b - коэффициент усиления адаптивного кодового словаря;

r - коэффициент масштабирования первого постпроцессорного средства;

$a_{улучш}$ - линейный множитель;

$f(b)$ - функция усиления b .

4. Синтезатор по п. 2 или 3, отличающийся

тем, что он выполнен с возможностью выделения коэффициента (r) масштабирования в соответствии с соотношением

$$b < \Pi_{нижн} r - 0,0$$

если

$$\Pi_{нижн} \leq b \leq \Pi_{верх} r - a_{улучш} b^2$$

$$b > \Pi_{верх} r - a_{улучш}$$

где $a_{улучш}$ - постоянная, которая управляет интенсивностью операции улучшения;

5. Синтезатор по п. 2, отличающийся тем, что он выполнен с возможностью выделения коэффициента (r) масштабирования первого постпроцессорного средства;

6. Синтезатор по п. 5, отличающийся тем, что в нем второй сигнал соответствует практически второму сигналу частичного возбуждения.

7. Синтезатор по любому из пп. 2-4, отличающийся тем, что в нем второй сигнал исходит из фиксированного кодового словаря.

8. Синтезатор по п. 7, отличающийся тем, что в нем второй сигнал соответствует практически первому сигналу частичного возбуждения.

9. Синтезатор по п. 7 или 8, отличающийся тем, что в нем средство регулировки усиления выполнено с возможностью масштабирования второго сигнала в соответствии с вторым коэффициентом (r') масштабирования, где

$$r' = -\frac{gr}{r+b},$$

где g - коэффициент масштабирования фиксированного кодового словаря;

b - коэффициент масштабирования адаптивного кодового словаря;

r - первый коэффициент масштабирования.

10. Синтезатор по любому из предыдущих пунктов, отличающийся тем, что в нем первый сигнал представляет собой первый сигнал возбуждения, пригодный для введения в речевой синтезирующий фильтр, а второй сигнал представляет собой второй сигнал возбуждения, пригодный для введения в речевой синтезирующий фильтр.

11. Синтезатор по любому из пп. 1-9, отличающийся тем, что в нем первый сигнал представляет собой первый речевой синтезированный сигнал, выходящий из первого речевого фильтра, а второй сигнал представляет собой выход из второго речевого синтезирующего фильтра.

12. Синтезатор по п. 11, отличающийся тем, что средство регулировки усиления выполнено с работой по сигналам, вводимым во второй речевой синтезирующий фильтр.

13. Синтезатор по любому из предыдущих пунктов, отличающийся тем, что он выполнен

для модификации первого сигнала путем объединения второго сигнала с первым сигналом.

14. Синтезатор по п. 13, отличающийся тем, что постпроцессорное средство дополнительно содержит средство адаптивной регулировки энергии, выполненное с возможностью масштабирования видоизмененного первого сигнала в соответствии со следующим соотношением:

$$k = \sqrt{\sum_{n=0}^{N-1} e_{x^2}(n) / \sum_{n=0}^{N-1} e_{w^2}(n)}$$

где N - соответственно выбранный период адаптации;

$e_{x^2}(n)$ - первый сигнал;

$e_{w^2}(n)$ - видоизмененный первый сигнал;

k - масштабный коэффициент энергии.

15. Способ улучшения синтезированной речи путем последующей обработки на основании коэффициентов линейного предсказания, включающий выделение первого сигнала, содержащего информацию о периодичности речи, из источника сигнала возбуждения, содержащего фиксированный кодовый словарь и адаптивный кодовый словарь, получение первого сигнала путем комбинирования первого и второго сигналов частичного возбуждения, исходящих соответственно из фиксированного и адаптивного кодовых словарей, видоизменение содержания информации о периодичности речи первого сигнала в соответствии с вторым сигналом, выделяемым из источника сигнала возбуждения путем масштабирования второго сигнала в соответствии с первым коэффициентом (p) масштабирования, выделенного из информации основного тона, связанной с первым сигналом, и комбинирования второго сигнала с первым сигналом.

16. Способ по п. 15, отличающийся тем, что первый коэффициент (p) масштабирования выделяют из коэффициента (b) усиления для информации основного тона первого сигнала.

17. Способ по п. 16, отличающийся тем, что первый коэффициент (p) масштабирования выделяют в соответствии со следующим уравнением:

$$b < \Pi_{\text{нижн}} p - 0,0$$

$$\Pi_{\text{нижн}} \leq b < \Pi_2 p - a_{\text{улучш1}} f_1(b)$$

$$\Pi_2 \leq b < \Pi_3 p - a_{\text{улучш2}} f_2(b)$$

если

$$\Pi_{N-1} \leq b \leq \Pi_{\text{верх}} p - a_{\text{улучшN-1}} f_{N-1}(b)$$

$$b > \Pi_{\text{верх}} p - a_{\text{улучшN}} f_N(b)$$

где Π - пороговые значения;

b - коэффициент усиления для информации основного тона первого сигнала; p - коэффициент масштабирования первого сигнала;

$a_{\text{улучш}}$ - линейный множитель;

$f(b)$ - функция b .

18. Способ по п. 16 или 17, отличающийся тем, что коэффициент (p) масштабирования выделяют в соответствии с соотношением

$$b < \Pi_{\text{нижн}} p - 0,0$$

если $\Pi_{\text{нижн}} \leq b \leq \Pi_{\text{верх}} p - a_{\text{улучш}} b^2$

$$b > \Pi_{\text{верх}} p - a_{\text{улучш}} b$$

где $a_{\text{улучш}}$ - постоянная, которая

управляет интенсивностью операции улучшения;

b - коэффициент усиления для информации основного тона первого сигнала;

Π - пороговые значения;

p - коэффициент масштабирования второго сигнала.

19. Способ по любому из пп. 15-18, отличающийся тем, что второй сигнал исходит из адаптивного кодового словаря.

20. Способ по п. 19, отличающийся тем, что второй сигнал соответствует практически второму сигналу частичного возбуждения.

21. Способ по любому из пп. 15-18, отличающийся тем, что второй сигнал исходит из фиксированного кодового словаря.

22. Способ по п. 21, отличающийся тем, что второй сигнал соответствует практически первому сигналу частичного возбуждения.

23. Способ по п. 21 или 22, отличающийся тем, что второй сигнал масштабируют в соответствии с вторым коэффициентом (p') масштабирования, где

$$p' = -\frac{gp}{p+b},$$

где g - коэффициент масштабирования фиксированного кодового словаря;

b - коэффициент масштабирования адаптивного кодового словаря;

p - первый коэффициент масштабирования.

24. Способ по любому из пп. 15-23, отличающийся тем, что в нем первый сигнал представляет собой первый сигнал возбуждения, пригодный для введения в первый речевой синтезирующий фильтр, а второй сигнал представляет собой второй сигнал возбуждения, пригодный для введения во второй речевой синтезирующий фильтр.

25. Способ по любому из пп. 15-23, отличающийся тем, что в нем первый сигнал представляет собой первый речевой синтезированный сигнал, выходящий из первого речевого фильтра, а второй сигнал представляет собой выход из второго речевого синтезирующего фильтра.

26. Способ по любому из пп. 15-25, отличающийся тем, что его осуществляют для видоизменения первого сигнала путем объединения второго сигнала с первым сигналом.

27. Способ по п. 26, отличающийся тем, что видоизмененный первый сигнал нормируют в соответствии со следующим соотношением:

$$k = \sqrt{\sum_{n=0}^{N-1} e_{x^2}(n) / \sum_{n=0}^{N-1} e_{w^2}(n)}$$

где N - соответственно выбранный период адаптации;

$e_{x^2}(n)$ - первый сигнал;

$e_{w^2}(n)$ - видоизмененный первый сигнал;

k - масштабный коэффициент энергии.

28. Радиоустройство, содержащее высокочастотное средство для приема радиосигнала и восстановления кодированной информации, содержащейся в этом радиосигнале, и синтезатор по любому из пп. 1-14.

29. Радиоустройство, выполненное с возможностью работы для улучшения синтезированной речи в соответствии со способом по любому из пп. 15-27.

30. Синтезатор для речевого синтеза, действующий на основании коэффициентов линейного предсказания и содержащий адаптивный кодовый словарь и фиксированный кодовый словарь для генерирования соответственно первого и второго сигналов частичного возбуждения, средство для масштабирования первого и второго сигналов частичного возбуждения соответственно с помощью полученных коэффициентов масштабирования адаптивного и фиксированного кодовых словарей, средство для видоизменения первого сигнала возбуждения в соответствии с еще одним коэффициентом масштабирования, причем коэффициент масштабирования является функцией информации основного тона, связанной с первым сигналом возбуждения, и средство для комбинирования второго сигнала частичного возбуждения с видоизмененным первым сигналом частичного возбуждения.

31. Синтезатор для речевого синтеза, действующий на основании коэффициентов линейного предсказания и содержащий адаптивный кодовый словарь и фиксированный кодовый словарь для генерирования соответственно первого и второго сигналов частичного возбуждения, средство для масштабирования первого и второго сигналов частичного возбуждения с помощью полученных коэффициентов масштабирования соответственно адаптивного и фиксированного кодовых словарей, средство для видоизменения второго сигнала возбуждения в соответствии с еще одним коэффициентом масштабирования, причем коэффициент масштабирования является функцией информации основного тона, связанной с первым сигналом возбуждения, и средство для комбинирования видоизмененного второго сигнала частичного возбуждения с первым сигналом частичного возбуждения.

32. Синтезатор по п. 30, отличающийся тем, что в нем первый коэффициент (а) масштабирования представлен в виде $a = b + p$, где b - усиление адаптивного кодового словаря, а p - коэффициент усиления улучшенного восприятия, выделяемый из следующего соотношения:

$$b < \Pi_{\text{нижн}} p - 0,0$$

$$\Pi_{\text{нижн}} \leq b < \Pi_2 p - a_{\text{улучш}} f_1(b)$$

$$\Pi_2 \leq b < \Pi_3 p - a_{\text{улучш}} f_2(b)$$

если

$$\Pi_{N-1} \leq b \leq \Pi_{\text{верх}} p - a_{\text{улучш}} f_{N-1}(b)$$

$$b > \Pi_{\text{верх}} p - a_{\text{улучш}} f_N(b)$$

где Π - пороговые значения;

b - коэффициент усиления адаптивного кодового словаря;

p - коэффициент усиления улучшенного восприятия;

$a_{\text{улучш}}$ - линейный множитель;

$f(b)$ - функция усиления b .

33. Синтезатор по п. 32, отличающийся тем, что он выполнен с обеспечением выделения коэффициента p усиления улучшенного восприятия в соответствии с соотношением

$$b < \Pi_{\text{нижн}} p - 0,0$$

если

$$\Pi_{\text{нижн}} \leq b \leq \Pi_{\text{верх}} p - a_{\text{улучш}} b^2$$

$$b > \Pi_{\text{верх}} p - a_{\text{улучш}}$$

и определяемого для p , представляющего

собой коэффициент усиления улучшенного восприятия.

34. Синтезатор по п. 31, отличающийся тем, что в нем второй коэффициент (a') масштабирования удовлетворяет следующему соотношению:

$$a' = -\frac{gp}{(p+b)},$$

где g - коэффициент масштабирования фиксированного кодового словаря;

b - коэффициент масштабирования адаптивного кодового словаря;

p - коэффициент усиления улучшенного восприятия, выделяемый в соответствии с соотношением

$$b < \Pi_{\text{нижн}} p - 0,0$$

$$\Pi_{\text{нижн}} \leq b < \Pi_2 p - a_{\text{улучш}} f_1(b)$$

$$\Pi_2 \leq b < \Pi_3 p - a_{\text{улучш}} f_2(b)$$

если

$$\Pi_{N-1} \leq b \leq \Pi_{\text{верх}} p - a_{\text{улучш}} f_{N-1}(b)$$

$$b > \Pi_{\text{верх}} p - a_{\text{улучш}} f_N(b)$$

где Π - пороговые значения;

b - коэффициент усиления адаптивного кодового словаря;

p - коэффициент усиления улучшенного восприятия;

$a_{\text{улучш}}$ - линейный множитель;

$f(b)$ - функция усиления b .

35. Синтезатор по п. 34, отличающийся тем, что он выполнен с возможностью выделения коэффициента p усиления улучшенного восприятия в соответствии с соотношением

$$b < \Pi_{\text{нижн}} p - 0,0$$

если

$$\Pi_{\text{нижн}} \leq b \leq \Pi_{\text{верх}} p - a_{\text{улучш}} b^2$$

$$b > \Pi_{\text{верх}} p - a_{\text{улучш}}$$

и определяемого для p , представляющего собой коэффициент усиления улучшенного восприятия.

36. Синтезатор по любому из пп. 30-35, отличающийся тем, что он выполнен с возможностью объединения первого и второго сигналов возбуждения после видоизменения.

37. Синтезатор по п. 36, отличающийся тем, что он дополнительно содержит адаптивное средство регулировки энергии для видоизменения объединенных масштабированных первого и второго сигналов в соответствии со следующим соотношением:

$$k = \sqrt{\frac{\sum_{n=0}^{N-1} ex^2(n)}{\sum_{n=0}^{N-1} ew'^2(n)}}$$

где N - соответственно выбранный период адаптации;

$ex(n)$ - объединенные первый и второй сигналы;

$ew'(n)$ - объединенные масштабированные первый и второй сигналы;

k - масштабный коэффициент энергии.

38. Способ речевого синтеза на основании коэффициентов линейного предсказания, включающий в себя выделение первого сигнала частичного возбуждения из адаптивного кодового словаря и второго сигнала частичного возбуждения из фиксированного кодового словаря, масштабирование первого и второго сигналов частичного возбуждения соответственно с помощью полученных коэффициентов масштабирования адаптивного и

фиксированного кодовых словарей, видоизменение первого сигнала частичного возбуждения в соответствии с еще одним коэффициентом масштабирования, причем коэффициент масштабирования является функцией информации основного тона, связанной с первым сигналом частичного возбуждения, и комбинирование второго сигнала частичного возбуждения с видоизмененным первым сигналом частичного возбуждения.

39. Способ речевого синтеза на основании коэффициентов линейного предсказания, включающий в себя выделение первого сигнала частичного возбуждения из адаптивного кодового словаря и второго сигнала частичного возбуждения из

фиксированного кодового словаря, масштабирование первого и второго сигналов частичного возбуждения соответственно с помощью полученных коэффициентов масштабирования адаптивного и фиксированного кодовых словарей, видоизменение второго сигнала частичного возбуждения в соответствии с еще одним коэффициентом масштабирования, причем коэффициент масштабирования является функцией информации основного тона, связанной с первым сигналом частичного возбуждения, и комбинирование видоизмененного второго сигнала частичного возбуждения с первым сигналом частичного возбуждения.

15

20

25

30

35

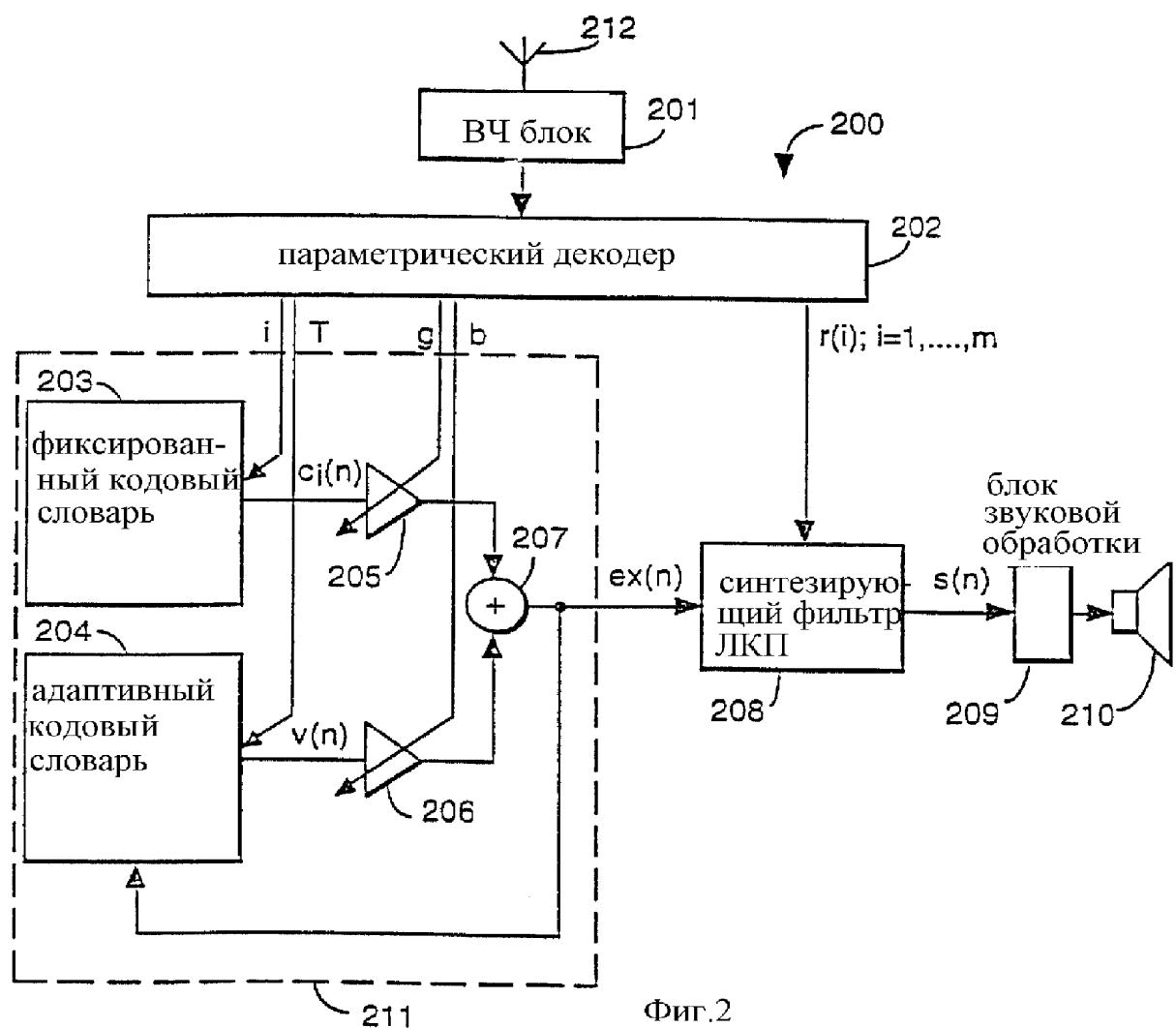
40

45

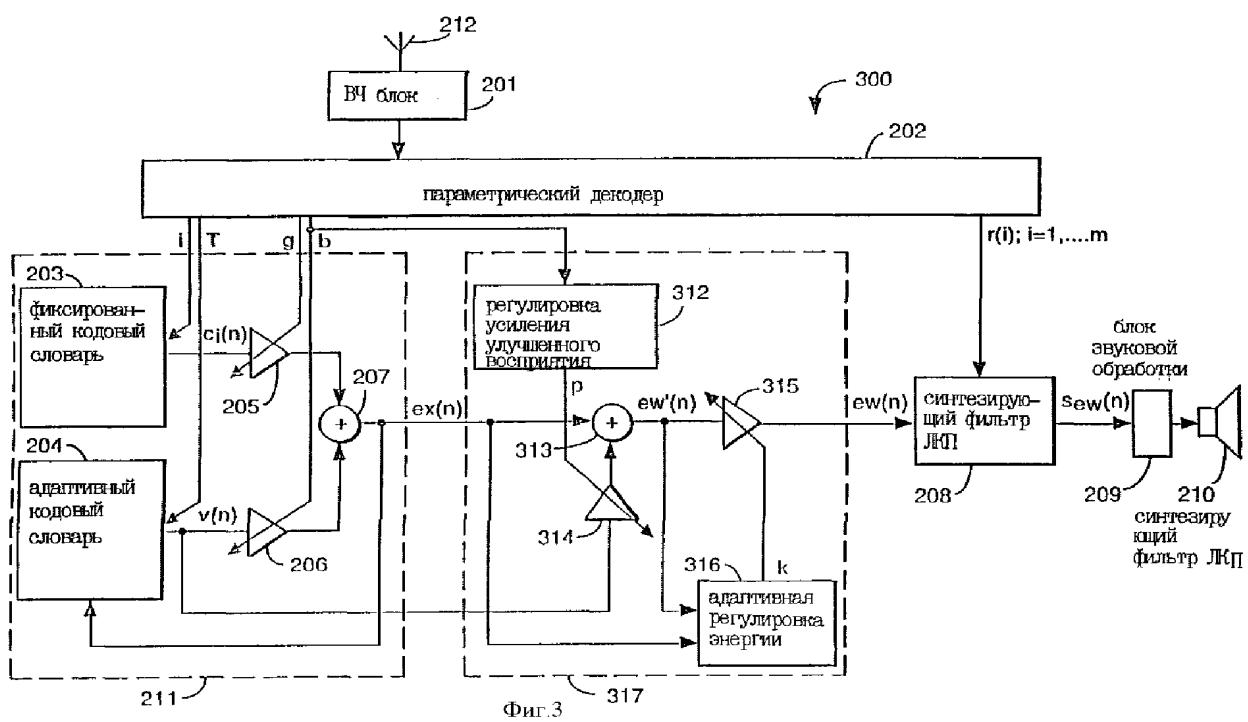
50

55

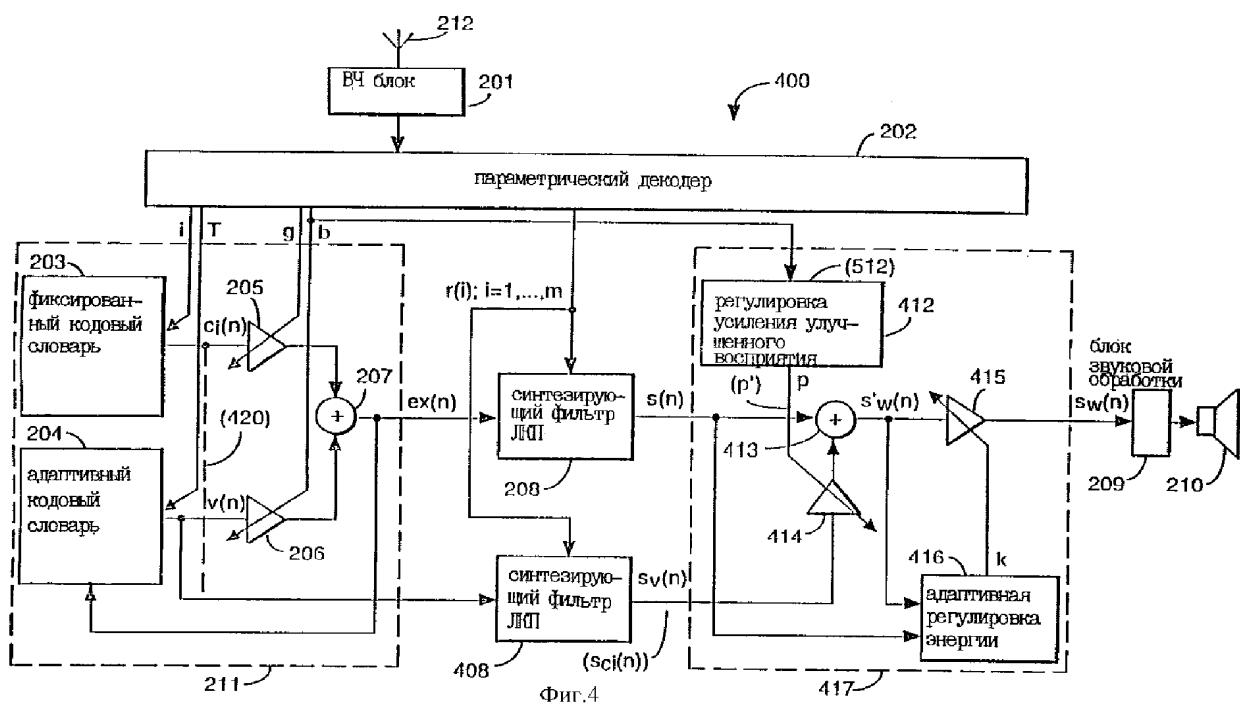
60



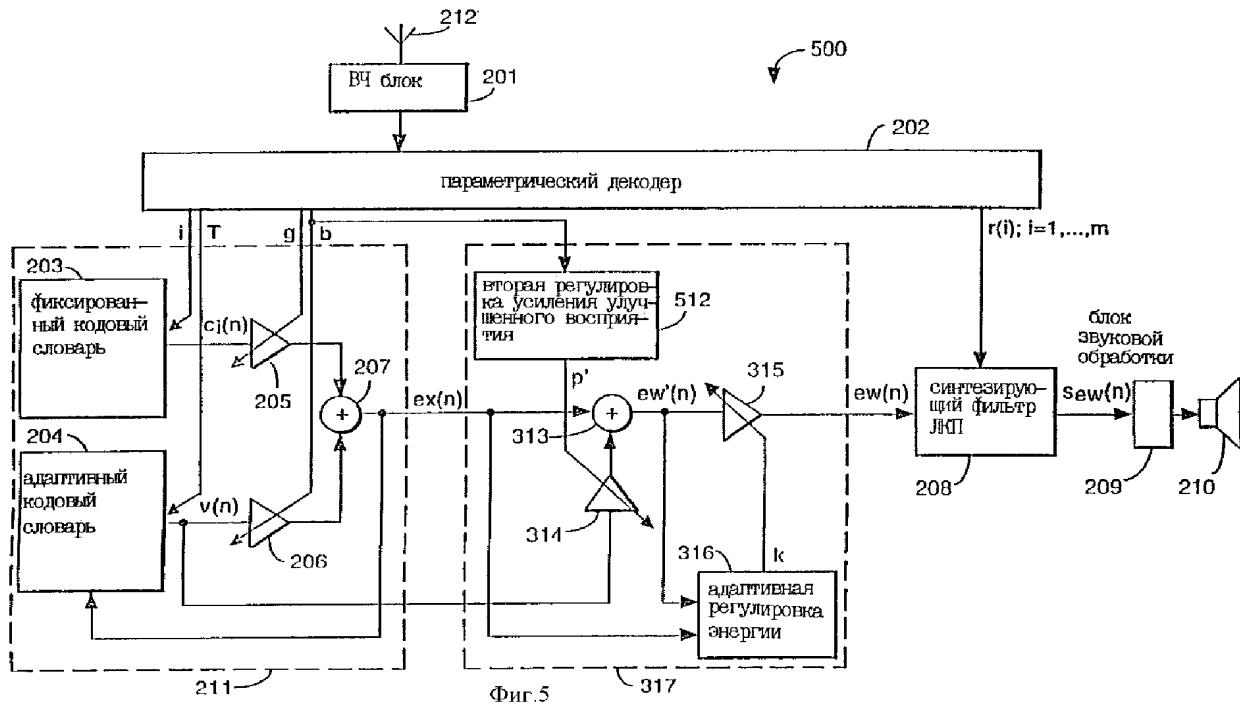
Фиг.2



Фиг.3



Фиг. 4



Фиг. 5

